Qualitative Likelihoodism

Abstract

We investigate whether qualitative analogs of the likelihood principle (LP) and related likelihoodist theses can be justified by informal, everyday reasoning. To do so, we first *formulate* qualitative analogs of two likelihoodist theses: LP and the law of likelihood (LL). Then, using a framework for qualitative conditional probability, we prove three novel theorems that suggest that, just as LP conflicts with virtually every non-Bayesian approach to statistical inference, the qualitative analogs of LP and LL motivate – and are best motivated by – qualitative forms of Bayesian reasoning.

The central question of this paper is: can qualitative analogs of the likelihood principle (LP) and related likelihoodist theses be justified by informal, everyday reasoning? LP has profound implications, as it conflicts with virtually all non-Bayesian approaches to statistical inference, including the use of p-values and confidence intervals (Berger and Wolpert 1988).

Although LP is formalized using probability theory, it is often stated in everyday English and motivated using intuitive concepts. For instance, (Birnbaum 1962, p. 271) says LP asserts the "irrelevance of [experimental] outcomes not actually observed." Related likelihoodist theses (e.g., the law of likelihood) extend heuristics for everyday reasoning, such as the rule that one should favor explanations that make one's data more probable (Royall 1997). Thus, it is natural to ask whether likelihoodist theses can be generalized and motivated in a qualitative, comparative framework, i.e., a framework in which one makes judgments like "A is more likely than B" without assigning events numerical probabilities.

Our central contributions in this paper are twofold. First, we formulate qualitative analogs of two likelihoodist theses: the law of likelihood (LL) and LP. This is straightforward for LL but tricky for LP. Second, using a framework for qualitative conditional probability, we prove two novel theorems that suggest that, just as LP conflicts with virtually every non-Bayesian approach to statistical inference, the qualitative analogs of LP and LL motivate – and are best motivated by – qualitative forms of Bayesian reasoning.

In sections one and two, we argue that likelihoodist theses characterize exactly when *all* Bayesian reasoners agree about how data ought to change one's posterior probabilities. To do so, we prove three simple probabilistic theorems that make precise the claim that LL and LP are "compatible" with Bayes' rule (Edwards 1984, p. 28).

Sections three and four contain most of our novel results. There we prove qualitative analogs of two of the theorems in section one using axioms for qualitative conditional probability (Krantz et al. 2006, p. 222-223). In other words, we model agents who must reason about statistical hypotheses but whose degrees of belief might not obey the axioms of quantitative probability.

Our results are important because they suggest that, in principle, some Bayesian and likelihood-based methods might be applicable even when one cannot justifiably use a precise, quantitative probability distribution to model (i) the likelihood of obtaining particular data or (ii) an experimenter's prior/posterior beliefs (or both) (Giang and Shenoy 2005). Our current results, however, are theoretical, and so our framework awaits application to data.

Our work is part of the larger project of modeling non-probabilistic reasoning under uncertainty (Halpern 2017); similar models include plausibility functions (Dubois and Prade 1998, 2012), belief functions (Shafer 1976, Dempster 2008), imprecise probabilities (Walley 1991), and ranking theory (Spohn 2012). Our model differs from the above frameworks in that we take *conditional* probability to be primitive, like (Keynes (2004), Popper (1959) and Renyi (2007). Our work is closest to (Smets 1993), who states a version of LP for plausibility functions, and to (Wasserman 1990), who explores the relationship between Dempster-Shafer theory and "Bayesian robustness", i.e., the practice of testing whether a statistical inference is valid for many prior distributions. We show that qualitative versions of LL and LP characterize *agreements* among agents with different qualitative "priors" that need not satisfy the standard probability axioms.

Likelihoodism

At least two distinct theses are called "the likelihood principle" and are used to motivate likelihood-based methods (e.g., maximum likelihood estimation) and Bayesian tools. For clarity, we distinguish the two:

- Law of likelihood (LL): $P_{\theta_1}(E) > P_{\theta_2}(E)$ if and only if the data E favors θ_1 over θ_2.
- Likelihood principle (LP): Let E and F are outcomes of two experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. If there is some $c > 0$ such that $P^{\mathbb{E}}_\theta(E) = c \cdot P^{\mathbb{F}}_\theta(F)$ for all $\theta \in \Theta$, then E and F are evidentially equivalent.

Here, Θ is the possible values of some unknown parameter. Elements of Θ are called *simple/point* hypotheses, and subsets are called *composite*. To avoid measurability assumptions, we typically assume composite hypotheses are finite.

Some defend a thesis stronger than LL, namely that the likelihood ratio $P_{\theta_1}(E)/P_{\theta_2}(E)$ quantifies the degree to which the data E favors θ_1 over θ_2. Call that thesis LL^+. What makes LL^+ strictly stronger than LL is that it allows one to compare how strong two pieces of evidence E and F are by assessing whether $P_{\theta_1}(E)/P_{\theta_2}(E) \geq P_{\theta_1}(F)/P_{\theta_2}(F)$ or vice versa.

One of the most common arguments for LL, LP, and LL^+ is that the principles are "compatible" with Bayes rule (Edwards 1984, p. 28). Here is how that "compatibility" is often explained for LL^+. Think of an experiment $\mathbb{E}$ as a pair $\langle \Omega^{\mathbb{E}}, \{P^{\mathbb{E}}_\theta\}_{\theta\in\Theta}\rangle$, where $\Omega^{\mathbb{E}}$ represents the possible outcomes of the experiment and $P^{\mathbb{E}}_\theta(\cdot)$ is a probability distribution (called a likelihood) over $\Omega^{\mathbb{E}}$ that specifies how likely each outcome is if θ is the true value of the parameter. Suppose Q is a prior probability distribution over Θ, and define posterior in the standard way:

$$Q^{\mathbb{E}}(H|E) = \frac{Q^{\mathbb{E}}(E \cap H)}{Q^{\mathbb{E}}(E)} := \frac{\sum_{\theta \in H} P^{\mathbb{E}}_\theta(E) \cdot Q(\theta)}{\sum_{\theta\in\Theta} P^{\mathbb{E}}_\theta(E) \cdot Q(\theta)}$$

for any $H \subseteq \Theta$. Then Bayes' Rule entails

$$\frac{Q^{\mathbb{E}}(\theta_1|E)}{Q^{\mathbb{E}}(\theta_2|E)} = \frac{P^{\mathbb{E}}_{\theta_1}(E)}{P^{\mathbb{E}}_{\theta_2}(E)} \cdot \frac{Q(\theta_1)}{Q(\theta_2)}$$

for any prior Q. So the likelihood ratio is a measure of the degree to which *all* Bayesians' posterior degrees of belief in θ_1 increase (or decrease) upon learning E. We introduce another way of capturing that same idea below.

We would like to suggest one new argument for LL^+, as it will be important below. LL^+ *unifies* LL and LP, in the sense that it entails both theses. Why? Assume that, for every piece of evidence E and any two hypotheses θ_1 and θ_2, there is a numerical degree $deg(E, \theta_1, \theta_2)$ to which E favors θ_1 over θ_2 (here, the degree might be negative).

To show LL^+ entails LL under plausible assumptions, say that E *favors* θ_1 over θ_2 if $deg(E, \theta_1, \theta_2) > deg(\Omega, \theta_1, \theta_2)$, in other words, if E provides better evidence for θ_1 (over θ_2) than the sure event. If LL^+ holds, then $deg(E, \theta_1, \theta_2) = P_{\theta_1}(E)/P_{\theta_2}(E)$, and so E favors θ_1 over θ_2 precisely if $P_{\theta_1}(E)/P_{\theta_2}(E) = deg(E, \theta_1, \theta_2) > deg(\Omega, \theta_1, \theta_2) = P_{\theta_1}(\Omega)/P_{\theta_2}(\Omega) = 1$. In other words, LL^+ plus the above assumptions entails that E favors θ_1 over θ_2 precisely if $P_{\theta_1}(E) > P_{\theta_2}(E)$, exactly as LL asserts.[1]

[1] Notice, we drop the superscript $\mathbb{E}$ when it's clear from context.

To show LL^+ entails LP under plausible assumptions, say that E and F are *evidentially equivalent* if $deg(E, \theta_1, \theta_2) = deg(F, \theta_1, \theta_2)$ for all θ_1 and θ_2. In other words, E and F favor all hypotheses by equal amounts. Again, if LL^+ holds, then $deg(E, \theta_1, \theta_2) = P_{\theta_1}(E)/P_{\theta_2}(E)$, and so E and F are evidentially equivalent precisely if

$$\frac{P_{\theta_1}(E)}{P_{\theta_2}(E)} = \frac{P_{\theta_1}(F)}{P_{\theta_2}(F)} \tag{1}$$

for all θ_1 and θ_2. If there is $c > 0$ such that $P_\theta(E) = c \cdot P_\theta(F)$ for all θ, then Equation 1 holds. So LL^+ entails LP.

In sum, (i) LL^+ is "compatible" with Bayes rule in the sense that the likelihood ratio is also the ratio of *all* Bayesians' posterior to prior probabilities, and (ii) LL^+ entails both LL and LP, which are also thought to be intuitively plausible (and are "compatible" with Bayesianism in still other ways). In the next section, we first prove that LL^+ is compatible with Bayes rule in another important way: it characterizes when *all* Bayesians agree that E is better evidence than F. In the second half of the paper, we show that the qualitative analog of such Bayesian agreement unifies qualitative analogs of LL and LP in the same way LL^+ unifies quantitative versions of those principles.

Bayesian Agreement and Likelihoodism

We now argue that likelihoodist theses characterize when all Bayesians agree about how evidence ought to change one's posterior. Before stating our main definition, we introduce some notation. Given an experiment $\mathbb{E}$, we let $\Delta^{\mathbb{E}} = \Theta \times \Omega^{\mathbb{E}}$. We use H_1, H_2 etc. to denote subsets of Θ, and E, F, etc. to denote subsets of Ω. We will write $Q(\cdot|H)$ and $Q(\cdot|E)$ instead of $Q(\cdot|H \times \Omega)$ and $Q(\cdot|\Theta \times E)$, and similarly for events to the left of the conditioning bar.

Definition 1. *Suppose $H_1, H_2 \subseteq \Theta$ are disjoint. Let E and F be outcomes of experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. Say E* Bayesian supports *H_1 over H_2* at least as much as *F if $Q^{\mathbb{E}}(H_1|E \cap (H_1 \cup H_2)) \geq Q^{\mathbb{F}}(H_1|F \cap (H_1 \cup H_2))$ for all priors Q for which those conditional probabilities are well-defined. If the inequality is strict for all such Q, then we say E* Bayesian supports *H_1 over H_2* strictly more than *F. In the former case, we write $E \;{}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$, and in the latter, we write $E \;{}_{H_1}\!\overset{\mathcal{B}}{\gg}_{H_2} F$.*

Our first goal is to find a necessary and sufficient condition for Bayesian support that involves only likelihoods. The following theorem does exactly that.

Theorem 1. *Suppose H_1 and H_2 are finite and disjoint. Let E and F be outcomes of experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. Then $E \;{}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$ if and only if for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$:*

$$P^{\mathbb{E}}_{\theta_1}(E) \cdot P^{\mathbb{F}}_{\theta_2}(F) \geq P^{\mathbb{E}}_{\theta_2}(E) \cdot P^{\mathbb{F}}_{\theta_1}(F) \tag{2}$$

Similarly, $E \;{}_{H_1}\!\overset{\mathcal{B}}{\gg}_{H_2} F$ if and only if the inequality in the above equation is strict.

The proof uses only basic probability theory, and so we omit it. The theorem asserts that $E \;{}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$ exactly when LL^+ says that E provides stronger evidence for θ_1 over θ_2 than F does. So, the Bayesian support relation is one way of clarifying what likelihoodist theses like LL$^+$ mean when they discuss "favoring." This is important because, as we show below, the Bayesian support relation has a direct qualitative analog, unlike LL$^+$, which deals explicitly with numerical degrees of favoring.

The Bayesian support relation is also important because it can be used to define the following notions of "Bayesian favoring" and "Bayesian favoring equivalence" that are respectively equivalent to the notions of "favoring" in LL and "evidential equivalence" in LP. This should be unsurprising given our argument above that LL$^+$ can be used to derive LL and LP under plausible assumptions.

Definition 2. *E Bayesian favors H_1 to H_2 if $E \ {}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} \Omega$. Say it strictly does if $E \ {}_{H_1}\!\overset{\mathcal{B}}{\gg}_{H_2} \Omega$.*

So Bayesian favoring is a special case of Bayesian support. In essence, the support relation *compares* (a) how much E Bayesian favors H_1 to H_2 to (b) how much F Bayesian favors H_1 to H_2 . Thus, the next theorem, which shows that Bayesian favoring is equivalent to LL, follows directly from Theorem 1.

Theorem 2. *E Bayesian favors H_1 to H_2 if and only if $P_{\theta_1}(E) \geq P_{\theta_2}(E)$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. If $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple, then E strictly Bayesian favors H_1 to H_2 if and only if* LL *entails E favors H_1 to H_2.*

We finish this section by studying two Bayesian notions of evidential equivalence that are (i) easy to translate to the qualitative setting and (ii) equivalent to LP.

Definition 3. *Say E and F are* Bayesian posterior equivalent *if $Q^{\mathbb{E}}(H|E) = Q^{\mathbb{F}}(H|F)$ for all hypotheses H and all priors Q for which at least one of those conditional probabilities is well-defined.*

Definition 4. *E and F are* Bayesian favoring equivalent *if $F \ {}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} E$ and $E \ {}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$ for all disjoint hypotheses H_1 and H_2.*

It is well-known that, if LP entails E and F are evidentially equivalent, then they are posterior equivalent; so Definition 3 makes precise how LP is compatible with Bayesianism. We show that Definition 4 is an equivalent notion, which is important because Definition 4 explains how LL$^+$ unifies LL and LP.

Theorem 3. *The following are equivalent: (1)* LP *entails E and F are evidentially equivalent, (2) E and F are Bayesian posterior equivalent, (3) E and F are Bayesian favoring equivalent.*

We provide a detailed proof in [Blinded for review]. The proof of $1 \Rightarrow 2$ is well-known. To our knowledge, no one has stated $2 \Rightarrow 1$, even though the proof is straightforward.

Qualitative Likelihoodism

Key Concepts

To move from quantitative to qualitative probability, we replace probability functions with two *orderings*. As before, let Θ be the set of simple hypotheses, and for any experiment $\mathbb{E}$, we let $\Omega^{\mathbb{E}}$ the set of experimental outcomes. The first relation, $\sqsubseteq$, is the qualitative analog of the set of likelihood functions. As before, we drop $\mathbb{E}$ when it's clear from context. Informally, $A|\theta \sqsubseteq B|\eta$ represents the claim that "experimental outcome B is at least as likely under supposition η as outcome A is under supposition θ"; it is the qualitative analog of $P_\theta(A) \leq P_\eta(B)$. We write $A|\theta \equiv B|\theta$ if $A|\theta \sqsubseteq B|\theta$ and vice versa.

Although we typically consider expressions of the form $A|\theta \sqsubseteq B|\eta$, the $\sqsubseteq$ relation is also defined when experimental outcomes $E \subseteq \Omega$ appear to the right of the conditioning bar, i.e., $A|\theta \cap E \sqsubseteq A|\theta \cap F$ is a well-defined expression if $E, F \subseteq \Omega$. However, it is not-well defined when composite hypotheses H appear to the right of the conditioning bar, just as there are no likelihood functions $P_H(\cdot)$ for composite hypotheses in the quantitative case.

Bayesians assume that beliefs are representable by a probability function. We weaken that assumption and assume beliefs are representable by an ordering $\preceq$ on $\mathcal{P}(\Delta) \times \mathcal{P}(\Delta)$, where $\mathcal{P}(\Delta)$ is the power set of Δ. As before, if $E \subseteq \Omega$ and $H \subseteq \Theta$ we write $E|H$ and instead of $\Theta \times E|H \times \Omega$. In the special case in which $H = \{\theta\}$ is a singleton, we omit the curly brackets and write $E|\theta$ instead of $E|\{\theta\}$.

The notation is suggestive; $A|B \preceq C|D$ if the experimenter regards C as at least as likely under supposition D as A would be under supposition B. Just as Bayesians condition on both composite hypotheses and experimental outcomes, there are no restrictions on what can appear to the right of the conditioning bar in expressions involving $\preceq$.

Main Results

It is now easy to state a qualitative analog of LL:

- Qualitative law of likelihood (QLL): E favors θ_1 to θ_2 if $E|\theta_2 \sqsubset E|\theta_1$.

To justify LL, we showed in Theorem 2 that LL characterizes precisely when all Bayesian agents agree E favors one hypothesis over another. By analogy, define:

Definition 5. *E qualitatively favors H_1 to H_2 if $H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$ for all orderings $\preceq$ satisfying the axioms below.*

Our first major result is the qualitative analog of Theorem 2. Just as Theorem 2 entails that E Bayesian favors θ_1 to θ_2 when LL entails so, our first major result shows that E qualitatively favors θ_1 to θ_2 if and only if QLL entails so.

Theorem 4. *Suppose H_1 and H_2 are finite. Then E qualitatively favors H_1 over H_2 if and only if $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. Further, the favoring inequality is strict exactly when the right inequality is strict. If $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple, then E qualitatively favors H_1 over H_2 if and only if* QLL *entails so.*

Our second main result is a qualitative analog of Theorem 3, which says that LP characterizes when two pieces of evidence are posterior and Bayesian-favoring equivalent. The qualitative analogs of posterior equivalence and favoring equivalence are obvious, but what about LP? For reasons of space, we state a qualitative analog of a restricted form of LP; the restricted form provides *sufficient* (but not necessary) conditions for posterior and favoring equivalence. To motivate the principle, notice that we can multiply probabilities when two events are independent. So suppose E and F are outcomes of the same experiment and

1. $P_\theta(E) = P_\theta(F \cap C_\theta)$ for all θ,
2. For all $\theta \in \Theta$, the events F and C_θ are conditionally independent given θ.
3. $P_\theta(C_\theta) = P_\upsilon(C_\upsilon) > 0$ for all $\theta, \upsilon \in \Theta$.

Roughly, the event C_θ acts as a witnesses to the equality $P_\theta(E) = c \cdot P_\theta(F)$. Specifically, assumptions 1 and 2 encode the equality, and assumption 3 asserts this constant is invariant with respect to the parameter θ.

By LP, the three conditions entail that E and F are evidentially equivalent. The proof is simple. Let $c = P_{\theta_0}(C_{\theta_0})$ for any $\theta_0 \in \Theta$. Then for all θ:

$$\begin{aligned} P^{\mathbb{E}}_\theta(E) &= P_\theta(F \cap C_\theta) \text{ by Assumption 1} \\ &= P^{\mathbb{E}}(F) \cdot P_\theta(C_\theta) \text{ by Assumption 2} \\ &= c \cdot P_\theta(F) \text{ by Assumption 3} \end{aligned}$$

Since $P_\theta(E) = c \cdot P_\theta(F)$ for all θ, then E and F are evidentially equivalent by LP.

So define LP^- to be the thesis that, if conditions 1-3 hold, then E and F are evidentially equivalent. We just showed that, if LP^- entails E and F are evidentially equivalent, then so does LP. By Theorem 3, it follows that E and F are also Bayesian posterior and favoring equivalent. Our second major result is the qualitative analog of that fact.

Because conditions 1-3 do not contain any arithmetic operations, each has a direct qualitative analog. The only slightly tricky condition is the second. To define a qualitative analog of conditional independence, note that events A and B are conditionally independent given C if and only if $P(A|B \cap C) = P(A|C)$. Analogously, we will say events A and B are qualitatively conditionally independent given C if $A|B \cap C \sim A|C$. In this case, we write $A \perp\!\!\!\perp_C B$.[2]

Our second main result is the following:

Theorem 5. *Let $\{C_\theta\}_{\theta \in \Theta}$ be events such that*

1. *$E|\theta \equiv F \cap C_\theta|\theta$ for all $\theta \in \Theta$,*
2. *$F \perp\!\!\!\perp_\theta C_\theta$ for all $\theta \in \Theta$ and all $\preceq \in \mathcal{C}$,*
3. *$C_\theta|\theta \equiv C_\eta|\eta$ for all $\theta, \eta \in \Theta$.*

If Θ is finite, then E and F are qualitatively posterior and favoring equivalent.

Axioms for Qualitative Probability

Here are some plausible axioms for conditional beliefs.

Axiom 0: $A|\theta \preceq B|\eta$ if and only if $A|\theta \sqsubseteq B|\eta$.

Axiom 1: $\preceq$ is a weak order (i.e., total, reflexive, and transitive),

Axiom 2: $A|A \sim B|B$ and $\Delta|\Delta \succeq A|B$,

– Here, $A|B \sim C|D$ if and only if $A|B \preceq C|D$ and vice versa.

Axiom 3: $A \cap B|B \sim A|B$,

Axiom 4: Suppose $A \cap B = A' \cap B' = \emptyset$. If $A|C \succeq A'|C'$ and $B|C \succeq B'|C'$, then $A \cup B|C \succeq A' \cup B'|C'$; moreover, if either hypothesis is $\succ$, then the conclusion is $\succ$.

Axiom 5: Suppose $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$. If $B|A \preceq C'|B'$ and $C|B \preceq B'|A'$, then $C|A \preceq C'|A'$; moreover, if either hypothesis is $\succ$, the conclusion is $\succ$.

Axiom 0 is our own; it asserts that conditional degrees of belief must match the likelihoods specified by an experiment. Axioms 1-5 are a subset of (Krantz et al. 2006)'s axioms for qualitative conditional probability. For several reasons (e.g., there is no Archimedean condition), our axioms are not sufficient for $\preceq$ to be representable as either a (conditional) probability measure or a set of probability measures. See (Alon and Lehrer 2014) for a recent representation theorem for sets of probability measures.

Axioms 1-4 are fairly analogous to facts of quantitative probability. Axiom 5 is useful because it allows us to "multiply" in a qualitative setting. To see its motivation, note that if $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$, then

$$P(C|B) = \frac{P(C)}{P(B)} \text{ and } P(B|A) = \frac{P(B)}{P(A)},$$

and similarly for the A', B', and C'. So if $P(B|A) \geq P(C'|B')$ and $P(C|B) \geq P(B'|A')$, then

$$\frac{P(B)}{P(A)} \geq \frac{P(C')}{P(B')} \text{ and } \frac{P(C)}{P(B)} \geq \frac{P(B')}{P(A')}.$$

When we multiply the left and right-hand sides of those inequalties, we obtain $P(C)/P(A) \geq P(C')/P(A')$, which is equivalent to $P(C|A) \geq P(C'|A')$ given our assumption about the nesting of the sets.

We sometimes use the following variant of Axiom 5:

Lemma 1. *Suppose $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$. If $B|A \preceq B'|A'$ and $C|B \preceq C'|B'$, then $C|A \preceq C'|A'$; moreover, if either hypothesis is $\succ$, the conclusion is $\succ$.*

Important Lemmas

We first state some lemmas. Some proofs are included to illustrate how the axioms work, but many proofs are excluded. Excluded proofs may be found in [blinded for review]. The first lemma is used repeatedly in most of our proofs. It describes a common, restricted application of Axiom 5.

Lemma 2. *Suppose $A \supseteq B \supseteq C$ and $A \supseteq B' \supseteq C$. If $B|A \succeq C|B'$, then $B'|A \succeq C|B$. Further, if the hypothesis is $\succ$, the conclusion is $\succ$.*

[2] Lemma 2 entails $A \perp\!\!\!\perp_C B$ if and only if $B \perp\!\!\!\perp_C A$.

Bayesian Statistics	**Qualitative Analog**
$P^{\mathbb{E}}_\theta(E) \leq P^{\mathbb{F}}_\upsilon(F)$	$E|\theta \sqsubseteq F|\upsilon$
$Q^{\mathbb{E}}(H_1|E) \leq Q^{\mathbb{F}}(H_2|F)$	$H_1|E \preceq H_2|F$
LL $\Leftrightarrow$ E Bayesian favors H_1 to H_2	QLL $\Leftrightarrow$ E qualitatively favors H_1 to H_2.
LP $\Leftrightarrow$ E and F are Bayesian posterior/favoring equivalent	QLP $\Leftrightarrow$ E and F are qualitatively posterior/favoring equivalent.

Proof. Assume $A \supseteq B \supseteq C$, $A \supseteq B' \supseteq C$, and $B|A \succeq C|B'$. Now suppose, for the sake of contradiction, that $B'|A \not\succeq C|B$. By Axiom 1 (specifically, the totality of $\preceq$), $C|B \succ B'|A$. Now, we apply Axiom 5 with $A' = A$, $C' = C$. We thus get $C|A \succ C'|A' = C|A$, which contradicts the reflexivity of the weak ordering (by Axiom 1). Thus, $B'|A \succeq C|B$. □

The next set of lemmas are all analogs of the following simple, quantitative fact: $P(E|H_1 \cup H_2)$ must be between $P(E|H_1)$ and $P(E|H_2)$.

Lemma 3. *Suppose $H_1 \cap H_2 = \emptyset$. Then, $E|H_1 \succeq E|H_2$ if and only if $E|H_1 \succeq E|(H_1 \cup H_2)$. Further, if either side of the biconditional is $\succ$, the other side is $\succ$ too.*

Lemma 4. *Suppose $H_1 \cap H_2 = \emptyset$. If $E|H_1 \sim E|H_2$ then $E|(H_1 \cup H_2) \sim E|H_1 \sim E|H_2$.*

Lemma 5. *Suppose $H_1 \cap H_2 = \emptyset$.*

1. *If $E|H_1, E|H_2 \preceq E|H_3$, then $E|(H_1 \cup H_2) \preceq E|H_3$.*
2. *If $E|H_3 \preceq E|H_1, E|H_2$, then $E|H_3 \preceq E|(H_1 \cup H_2)$.*

If the premise is not strict AND $E|H_1 \sim E|H_2$, then the conclusion is not strict. Otherwise, the conclusion is strict.

The following lemma is trivial but necessary to state.

Lemma 6. *$\emptyset|A \sim \emptyset|B$ for all A and B.*

The next lemma is an analogue of the definition of conditional probability, as $A|\Delta$ is essentially the unconditional qualitative probability of A.

Lemma 7. *$X|Y \preceq W|Y$ iff $X \cap Y|\Delta \preceq W \cap Y|\Delta$.*

Qualitative Law of Likelihood

Our goal in this section is to prove Theorem 4. The theorem follows immediately from following two propositions:

Proposition 1. *Suppose $H_1 \cap H_2 = \emptyset$. Then $E|H_1 \succeq E|H_2$ if and only if $H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. Further, if either side of the biconditional contains a strict inequality $\succ$, then so does the other.*

Proposition 2. *Suppose H_1 and H_2 are finite and that $H_1 \cap H_2 = \emptyset$. Then $E|H_2 \preceq E|H_1$ for all orderings $\preceq$ (satisfying the axioms above) if and only if $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. Further, the lefthand side of the biconditional contains the strict inequality $\prec$ if and only if the right-hand side contains the strict relation $\sqsubset$.*

Proof of Proposition 1: First we prove the left-to-right direction. Suppose $E|H_1 \succeq E|H_2$, where $H_1 \cap H_2 = \emptyset$. By Lemma 3, it follows that $E|H_1 \succeq E|(H_1 \cup H_2)$. Applying Axiom 3, we get

$$E \cap H_1|H_1 \succeq E \cap (H_1 \cup H_2)|H_1 \cup H_2 \tag{3}$$

Define:

$$\begin{aligned} &A = H_1 \cup H_2 \\ &B = H_1 \qquad\qquad B' = E \cap (H_1 \cup H_2) \\ &C = E \cap H_1 \end{aligned}$$

Note that equation (3) says that $C|B \succeq B'|A$. So, using the contrapositive of Lemma 2 tells us that $C|B' \succeq B|A$, i.e.,

$$E \cap H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$$

Applying Axiom 3 to the left-hand side of that inequality yields the desired result. Note that if we had assumed $E|H_1 \succ E|H_2$, then our conclusion would contain $\succ$ because both Lemma 2 and Lemma 3 yield strict comparisons.

In the right to left direction, suppose $H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. By Axiom 3:

$$\begin{aligned} &H_1|E \cap (H_1 \cup H_2) \\ &\quad \sim H_1 \cap E \cap (H_1 \cup H_2)|E \cap (H_1 \cup H_2) \\ &\quad \sim E \cap H_1|E \cap (H_1 \cup H_2) \end{aligned}$$

So, we know $E \cap H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. Now, we apply the contrapositive of Lemma 2 with:

$$\begin{aligned} &A = H_1 \cup H_2 \\ &B = E \cap (H_1 \cup H_2) \qquad B' = H_1 \\ &C = E \cap H_1 \end{aligned}$$

The containment relations are satisfied, and $C|B \succeq B'|A$. Thus, we get $C|B' \succeq B|A$, i.e.,

$$E \cap H_1|H_1 \succeq E \cap (H_1 \cup H_2)|(H_1 \cup H_2).$$

Applying Axiom 3 to both sides of the last inequality yields $E|H_1 \succeq E|H_1 \cup H_2$. So by Lemma 3, we get $E|H_1 \succeq E|H_2$, as desired. As before, if the premise were $\succ$, the conclusion would also be $\succ$ because the necessary lemmas would yield strict comparisons. □

Proof of Proposition 2: In the right-to-left direction, suppose $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. We want to show $E|H_2 \preceq E|H_1$ for all orderings $\preceq$. So let $\preceq$ be any ordering satisfying the axioms.

We show $E|H_2 \preceq E|H_1$ by induction on the maximum of the number of elements in H_1 or H_2.

In the base case, suppose $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ both have one element. Then by Axiom 0, it immediately follows that $E|H_2 \preceq E|H_1$.

For the inductive step, suppose the result holds for all natural numbers $m \leq n$, and assume that $H_1 = \{\theta_{1,1}, \ldots \theta_{1,k}\}$ and $H_2 = \{\theta_{2,1} \ldots, \theta_{2,l}\}$ where either k or l (or both) is equal to $n+1$. Define $H_1' = \{\theta_{1,1}\}$ and $H_1'' = H_1 \setminus H_1'$, and similarly, define $H_2' = \{\theta_{2,1}\}$ and $H_2'' = H_2 \setminus H_2'$. Then H_1', H_1', H_2', and H_2'' all have n or fewer elements, and by assumption, $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_2 \in H_2', H_2''$ and all $\theta_1 \in H_1', H_1''$. By inductive hypothesis, it follows that $E|H_1', E|H_1'' \preceq E|H_2', E|H_2''$. By repeated application of Lemma 5, it follows that $E|H_1 = E|(H_1' \cup H_1'') \preceq E|(H_2' \cup H_2'') = E|H_2$.

Note that if the premise were $\sqsubset$, Axiom 0 would prove the strict version of the base case. In the inductive step, every inequality would also be strict because Lemma 5 preserves the strictness. Thus, we would get $E|H_2 \prec E|H_1$.

In the left-to-right direction, suppose $E|H_2 \preceq E|H_1$ for all orderings $\preceq$. We want to show that $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. So fix $\theta_1 \in H_1$ and $\theta_2 \in H_2$. We must show $E|\theta_2 \sqsubseteq E|\theta_1$. Suppose for the sake of contradiction that $E|\theta_2 \not\sqsubseteq E|\theta_1$. We must find at least one ordering $\preceq$ that (i) satisfies Axioms 0-5 and (ii) entails $E|H_2 \not\preceq E|H_1$.

Let $\preceq$ be any ordering satisfying axioms 0-5 such that

$$\begin{aligned} \theta_1|\Delta, \theta_2|\Delta &\succ \emptyset|\Delta \\ A|\Delta &\sim \emptyset|\Delta \text{ if } \theta_1, \theta_2 \notin A \end{aligned}$$

Those two equations together say the qualitative "prior" $\preceq$ assigns "positive" probability to θ_1 and θ_2 and "zero" probability to all other hypotheses.

We now need the following two lemmas. The first says that $\{\theta_1, \theta_2\}$ as defined is the analog of a probability one event. The second is the qualitative analog of the fact that if A is a probability one event, then it can be intersected with either side of the conditioning bar without changing the relevant probability. Proofs can be found in [blinded for review].

Lemma 8. *Suppose $A|\Delta \sim \emptyset|\Delta$ for all sets A that don't contain θ_1 or θ_2. Then, $\{\theta_1, \theta_2\}|B \sim \{\theta_1, \theta_2\}|\{\theta_1, \theta_2\}$ for all B.*

Lemma 9. *Let A be an event such that $A|A \sim A|B$ for all B. Then for all C and all B*

1. $C|B \sim (A \cap C)|B$
2. $C|B \sim C|(B \cap A)$

Now, we apply Lemma 8 and then Lemma 9. By the second conclusion of Lemma 9, $E|H_1 \sim E|H_1 \cap \{\theta_1, \theta_2\} \sim E|\theta_1$, as H_1 contains θ_1 but not θ_2. Similarly, $E|H_2 \sim E|\theta_2$. Recall, we assumed for the sake of contradiction that $E|\theta_1 \not\preceq E|\theta_2$. So Axiom 0 and Axiom 1 entail $E|\theta_2 \prec E|\theta_1$. Thus, we get $E|H_1 \sim E|\theta_1 \prec E|\theta_2 \sim E|H_2$. By transitivity, $E|H_1 \prec E|H_2$, contradicting our assumption.

If the premise of this proposition contained a strict inequality $\prec$, the conclusion would contain $\sqsubset$, by swapping all strict and weak inequalities in the preceding argument.

□

A Qualitative Likelihood Principle

In this section, we prove Theorem 5. Let $X_\theta = F \cap C_\theta \cap \theta|\bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$. We show that $\theta|E \sim X_\theta$ and $\theta|F \sim X_\theta$ for all θ. Hence, $\theta|E \sim \theta|F$ for all θ. Using finite additivity, we conclude $H|E \sim H|F$ for all finite hypotheses H. So E and F are qualitatively posterior equivalent if Θ is finite.

We begin with two general lemmas, which we use to prove the $\theta|E \sim X_\theta$. Below, let $[n] = \{1, 2, \ldots, n\}$ be the first n natural numbers.

Lemma 10. *Suppose $B_1, \ldots, B_n$ partition G. If $A_i|B_i \sim C_i|B_i$ for all $i \in [n]$, then*

$$\bigcup_{i \in [n]} A_i \cap B_i|G \sim \bigcup_{i \in [n]} C_i \cap B_i|G \textit{ for all } i \in [n]$$

Proof. We first use Lemma 7 on the premise, which yields $A_i \cap B_i|\Delta \sim C_i \cap B_i|\Delta$ for all $i \in [n]$. Applying Axiom 4 n times yields $\bigcup_{i \in [n]} A_i \cap B_i|\Delta \sim \bigcup_{i \in [n]} C_i \cap B_i|\Delta$. Because $B_1, \ldots, B_n$ partition G, this is equivalent to $G \cap \bigcup_{i \in [n]} A_i \cap B_i|\Delta \sim G \cap \bigcup_{i \in [n]} C_i \cap B_i|\Delta$. Finally, we apply Lemma 7 again to get the desired result. □

The following lemma outlines conditions under which we can "add" two qualitative equations by unioning the conditioning events. We will use this in the proof of Lemma 12.

Lemma 11. *Suppose $Y \cap Y' = Z \cap Z' = \emptyset$, that $X|Y \sim W|Z$ and $X|Y' \sim W|Z'$. Finally, assume that $X|Y \sim \emptyset|\Delta$ (so $W|Z \sim \emptyset|\Delta$) or $X|Y' \sim \emptyset|\Delta$ (so $W|Z' \sim \emptyset|\Delta$), or both. Then $X|Y \cup Y' \sim W|Z \cup Z'$ if and only if $Y|Y \cup Y' \sim Z|Z \cup Z'$.*

Lemma 12. *Suppose $B_1, \ldots, B_n$ partition G. Further, suppose $A|B_i \sim C_i|B_i$ for all $i \in [n]$. Then, $B_i|A \cap G \sim B_i \cap C_i|\bigcup_{j \in [n]} B_j \cap C_j$ for all $i \in [n]$.*

Lemma 12 can be thought of as an application of Bayes' theorem combined with the Law of Total Probability. If $C_i = C$ for all i, then quantitative analog of the lemma asserts the following: if $B_1 \ldots B_n$ partition G and $P(A|B_i) = P(C|B_i)$ for all i, then $P(B_i|A \cap G) = P(B_i|C \cap G)$ for all i. That claim follows easily from an application of Bayes' theorem and the Law of Total Probability.

Proof. By induction on n. In the base case, when $n = 1$, $B_1 = G$. Then:

$$\begin{aligned} G|A \cap G &\sim A \cap G|A \cap G && \text{by Axiom 3} \\ &\sim C_1 \cap G|C_1 \cap G && \text{by Axiom 2} \end{aligned}$$

which is the desired equality.

For the inductive step, assume $B_1, \ldots, B_{n+1}$ partition G and $A|B_i \sim C_i|B_i$ for all $i \in [n+1]$. Let $G' = G \setminus B_{n+1}$. Clearly, $B_1, \ldots, B_n$ partition G' and $A|B_i \sim C_i|B_i$ for all $i \in [n]$. Thus, by the inductive hypothesis:

$$B_i|A \cap G' \sim B_i \cap C_i|\bigcup_{j \in [n]} B_j \cap C_j \text{ for all } i \in [n] \tag{4}$$

Equation 4 also holds for $i = n+1$ because both $B_{n+1} \cap (A \cap G')$ and $(B_{n+1} \cap C_{n+1}) \cap \bigcup_{j \in [n]} B_j \cap C_j$ are empty. Thus, we can use Axiom 3 and Lemma 6 to get Equation 4. Further, it's easy to show that for all $i \in [n+1]$:

$$B_i|A \cap B_{n+1} \sim B_i \cap C_i|B_{n+1} \cap C_{n+1} \tag{5}$$

If $i = n+1$, one can simply use Axiom 3 to turn the left side into the sure event and then use Axiom 2, just like in the base case. If $i \neq n+1$, then note that $B_i \cap (A \cap B_{n+1}) = (B_i \cap C_i) \cap (B_i \cap C_{n+1}) = \emptyset$. So like above, we use Axiom 3 and Lemma 6 to get the equality. Our goal is to show:

$$B_i|A \cap G \sim B_i \cap C_i|\bigcup_{j \in [n+1]} B_j \cap C_j \text{ for all } i \in [n+1] \tag{6}$$

Equation 6 would follow if we could union the conditioning events in Equation 4 and Equation 5. Lemma 11 was designed for that purpose. Let i be arbitrary and define:

$$\begin{aligned} X &= B_i & W &= B_i \cap C_i \\ Y &= A \cap B_{n+1} & Z &= B_{n+1} \cap C_{n+1} \\ Y' &= A \cap G & Z' &= \bigcup_{k \leq n} B_k \cap C_k \end{aligned}$$

We want to apply Lemma 11, so we first check all the conditions. Note that $Y \cap Y' = Z \cap Z' = \emptyset$, since the B_i's are disjoint. Further, $X|Y \sim W|Z$ is simply Equation 5 and $X|Y' \sim W|Z'$ is Equation 4. Lastly, if $i = n+1$, $X|Y' = B_i|A \cap G' \sim \emptyset|\Delta$ by the disjointness of B_{n+1} and G', along with Axiom 3, and Lemma 6. If $i \leq n$, $X|Y = B_i|A \cap B_{n+1} \sim \emptyset|\Delta$, again because B_i and B_{n+1} are disjoint. So either $X|Y \sim \emptyset|\Delta$ or $X|Y' \sim \emptyset|\Delta$. With the conditions satisfied, Lemma 11 tells us that Equation 6 holds if and only if $Y|Y \cup Y' \sim Z|Z \cup Z'$ i.e.:

$$A \cap B_{n+1}|A \cap G \sim B_{n+1} \cap C_{n+1}| \bigcup_{k \leq n+1} B_k \cap C_k \quad (7)$$

To prove Equation 7, we first show:

$$A \cap B_{n+1}|G \sim B_{n+1} \cap C_{n+1}|G \quad (8)$$

Define:

$$\begin{aligned} X &= G & X' &= G \\ Y &= B_{n+1} & Y' &= B_{n+1} \\ Z &= A \cap B_{n+1} & Z' &= B_{n+1} \cap C_{n+1} \end{aligned}$$

Clearly, $Y|X \sim Y'|X'$, and $Z|Y = A \cap B_{n+1}|B_{n+1} \sim A|B_{n+1}$ by Axiom 3. Similarly, $Z'|Y' \sim C_{n+1}|B_{n+1}$. By the assumption that $A|B_i \sim C_i|B_i$ for all $i \in [n+1]$, we get $Z|Y \sim Z'|Y'$. So Lemma 1 tells us that $Z|X \sim Z'|X'$, which is Equation 8.

We now prove Equation 7 by contradiction. Suppose $A \cap B_{n+1}|A \cap G \not\sim B_{n+1} \cap C_{n+1}| \bigcup_{k \leq n+1} B_k \cap C_k$, and without loss of generality, say $A \cap B_{n+1}|A \cap G \succ B_{n+1} \cap C_{n+1}| \bigcup_{k \leq n+1} B_k \cap C_k$. Now, define:

$$\begin{aligned} X &= G & X' &= G \\ Y &= A \cap G & Y' &= \bigcup_{k \leq n+1} B_k \cap C_k \\ Z &= A \cap B_{n+1} & Z' &= B_{n+1} \cap C_{n+1} \end{aligned}$$

By Lemma 10, $Y|X \sim Y'|X' = A \cap G|G \sim \bigcup_{k \leq n+1} B_k \cap C_k|G$. Note that $Z|Y \succ Z'|Y'$ by assumption. So Lemma 1 entails that $Z|X \succ Z'|X'$, contradicting Equation 8. □

Proposition 3. *Suppose $E|\theta \equiv F \cap C_\theta|\theta$ for all $\theta \in \Theta$. If Θ is finite, then $\theta|E \sim F \cap C_\theta \cap \theta| \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$ for all $\theta \in \Theta$.*

Proof. We apply Lemma 12. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ and $B_i = \theta_i$. Thus, $G = \Theta$, which, recall, is shorthand for Δ in this context. Let $A = E$ and $C_i = F \cap C_{\theta_i}$. Then, $A|B_i \sim C_i|B_i$ is the premise of this lemma, and $B_i|A \sim B_i \cap C_i| \bigcup_{j \in [n]} B_j \cap C_j$ is the conclusion. □

Lemma 13. *Suppose $B_1, \ldots, B_n$ partition G. Further, suppose $A \perp\!\!\!\perp_{B_i} C_i$ and for all $i, j \in [n] : C_i|B_i \sim C_j|B_j$. Then, $\bigcup_{i \in [n]} C_i \cap B_i|A \cap G \sim C_k|B_k$, for all $k \in [n]$.*

The proof is by induction on the size of G. See [Blinded for review] for details.

Proposition 4. *Suppose $F \perp\!\!\!\perp_\theta C_\theta$ and $C_\theta|\theta \equiv C_\eta|\eta$ for all $\theta, \eta \in \Theta$. If Θ is finite, then $\theta|F \sim F \cap C_\theta \cap \theta| \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$ for all $\theta \in \Theta$.*

Proof. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ and let $B_i = \{\theta_i\}$. Thus, $G = \Theta$. Let $A = F$ and $C_i = C_\theta$. Then, from Lemma 13 we get $\bigcup_{\eta \in \Theta} C_\eta \cap \eta|F \sim C_\theta|\theta$ for all $\theta \in \Theta$. Now, since $F \perp\!\!\!\perp_\theta C_\theta$, we can change this to:

$$\bigcup_{\eta \in \Theta} C_\eta \cap \eta|F \sim C_\theta|F \cap \theta \quad (9)$$

Now, we apply Lemma 2 with:

$$\begin{aligned} &A = F \\ &B = \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta \qquad B' = F \cap \theta \\ &C = F \cap C_\theta \cap \theta \end{aligned}$$

Equation 9 says $B|A \sim C|B'$. So by Lemma 2, we get $B'|A \sim C|B$, which is the desired conclusion. □

We have shown that the three conditions of Theorem 5 entail that E and F are qualitatively posterior equivalent if Θ is finite. To show those conditions entail that E and F are qualitatively favoring equivalent, we need the following theorem, which is proven in a way exactly analogous to the proof of 2 ⇔ 3 of Theorem 3.

Theorem 6. *E and F are qualitatively posterior equivalent if and only if they are qualitatively favoring equivalent.*

Conclusion and Future Work

We have shown that, just as LL and LP characterize when all Bayesian reasoners agree how evidence should alter one's posterior probabilities, qualitative likelihoodist theses like QLL characterize agreement among agents whose beliefs satisfy fairly weak coherence requirements.

Our work can be extended in a number of ways. First and most importantly, our qualitative framework should be applied to (i) real data, (ii) common statistical models, and (iii) classification problems. See (Come et al. 2009), (Denoeux 2014), and (Denoeux 1995) respectively for work analogous to i-iii on Dempster-Shafer functions.

Second, one should characterize necessary conditions for qualitative posterior equivalence; the qualitative analog of LP^- that we state is only sufficient. Luckily, the *conditionality principle*, which is equivalent to LP quantitatively (Evans, Fraser, and Monette 1986), has a direct qualitative analog. We conjecture that, using Theorem 5, one can prove that the qualitative analog of the conditionality principle provides necessary and sufficient conditions for qualitative posterior and favoring equivalence.

Third, our qualitative theorems assume either that hypotheses are finite (as Theorem 4 does) or that the parameter space is finite (as Theorem 5). Assuming a qualitative countable additivity assumption (Krantz et al. 2006, p. 216), we conjecture, will allow us to extend the first result to countable hypotheses, but qualitative analogs of certain measure-theoretic results will be necessary for uncountable spaces.

References

[Alon and Lehrer 2014] Alon, S., and Lehrer, E. 2014. Subjective multi-prior probability: A representation of a partial likelihood relation. *Journal of Economic Theory* 151:476–492.

[Berger and Wolpert 1988] Berger, J. O., and Wolpert, R. L. 1988. The likelihood principle. In *Ims Lecture Notes-Monograph*, volume 6. Institute of Mathematical Statistics.

[Birnbaum 1962] Birnbaum, A. 1962. On the foundations of statistical inference. *Journal of the American Statistical Association* 57(298):269–306.

[Come et al. 2009] Come, E.; Oukhellou, L.; Denoeux, T.; and Aknin, P. 2009. Learning from partially supervised data using mixture models and belief functions. *Pattern recognition* 42(3):334–348.

[Dempster 2008] Dempster, A. P. 2008. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer. 57–72.

[Denoeux 1995] Denoeux, T. 1995. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* 25(5):804–813.

[Denoeux 2014] Denoeux, T. 2014. Likelihood-based belief function: justification and some extensions to low-quality data. *International Journal of Approximate Reasoning* 55(7):1535–1547.

[Dubois and Prade 1998] Dubois, D., and Prade, H. 1998. Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*. Springer. 169–226.

[Edwards 1984] Edwards, A. W. F. 1984. *Likelihood*. CUP Archive.

[Evans, Fraser, and Monette 1986] Evans, M. J.; Fraser, D. A.; and Monette, G. 1986. On principles and arguments to likelihood. *Canadian Journal of Statistics* 14(3):181–194.

[Giang and Shenoy 2005] Giang, P. H., and Shenoy, P. P. 2005. Decision making on the sole basis of statistical likelihood. *Artificial Intelligence* 165(2):137–163.

[Halpern 2017] Halpern, J. Y. 2017. *Reasoning about uncertainty*. MIT press.

[Keynes 2004] Keynes, J. M. 2004. *A Treatise on Probability*. Courier Dover Publications.

[Krantz et al. 2006] Krantz, D. H.; Luce, R. D.; Suppes, P.; and Tversky, A. 2006. *Foundations of Measurement Volume I: Additive and Polynomial Representations*. Mineola, NY: Dover Publications.

[Popper 1959] Popper, K. R. 1959. The propensity interpretation of probability. *The British journal for the philosophy of science* 10(37):25–42.

[Renyi 2007] Renyi, A. 2007. *Foundations of probability*. Courier Corporation.

[Royall 1997] Royall, R. 1997. *Statistical evidence: a likelihood paradigm*, volume 71. Chapman & Hall/CRC.

[Shafer 1976] Shafer, G. 1976. *A mathematical theory of evidence*, volume 42. Princeton university press.

[Smets 1993] Smets, P. 1993. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of approximate reasoning* 9(1):1–35.

[Spohn 2012] Spohn, W. 2012. *The laws of belief: Ranking theory and its philosophical applications*. Oxford University Press.

[Walley 1991] Walley, P. 1991. *Statistical reasoning with imprecise probabilities*. Number 42 in Monographs on Statistics and Applied Probability. New York: Chapman & Hall.

[Wasserman 1990] Wasserman, L. A. 1990. Belief functions and statistical inference. *Canadian Journal of Statistics* 18(3):183–196.